

Hard Prompts Made Easy: Gradient-Based Discrete Optimization for Prompt Tuning and Discovery

Year: 2023 | Citations: 346 | Authors: Yuxin Wen, Neel Jain, John Kirchenbauer, Micah Goldblum, Jonas Geiping

Abstract

The strength of modern generative models lies in their ability to be controlled through text-based prompts. Typical "hard" prompts are made from interpretable words and tokens, and must be hand-crafted by humans. There are also "soft" prompts, which consist of continuous feature vectors. These can be discovered using powerful optimization methods, but they cannot be easily interpreted, re-used across models, or plugged into a text-based interface. We describe an approach to robustly optimize hard text prompts through efficient gradient-based optimization. Our approach automatically generates hard text-based prompts for both text-to-image and text-to-text applications. In the text-to-image setting, the method creates hard prompts for diffusion models, allowing API users to easily generate, discover, and mix and match image concepts without prior knowledge on how to prompt the model. In the text-to-text setting, we show that hard prompts can be automatically discovered that are effective in tuning LMs for classification.